Incentives, Business Model and Required Technical Infrastructure for the Economic Aware Data Grid

Dang Minh Quan[1,*], Orestis Terzidis[2], and Sonja Klingert[1]

[1] School of Information Technology, International University in Germany, Germany
[2] SAP research, Germany

Abstract. The data Grids appear mostly in the scientific community but it is also necessary to have the data Grid in the industry sector. Applying the way of scientific data Grid to the industry brings many disadvantages such as expensive investment, not efficient usage and unfair sharing. This paper presents an alternative approach, an economic aware data Grid. The incentives of having an economic aware data Grid, the business model and the working mechanism are analyzed in depth.

1 Introduction

The data Grid appears from the requirement of sharing, accessing, transporting, processing, and managing large data collections distributed worldwide. At the present, the data Grids appear mostly in the scientific community [1]. A data Grid usually includes many High Performance Computing Centers (HPCCs) with enormous storage capacity and computing power. It is obvious that investing a data Grid need a lot of money. The financial sources for building such data Grid are from governments and scientific funding foundations. For that reason, researchers within the data Grid can use those resources freely.

In the industry sector, the data Grid is also necessary. We can look at the scenario of investment banks. The investment bank has many branches distributed over the geography. Each branch has its own business. Each branch usually runs data mining applications over the set of collected financial data. With the time, this activity becomes important and need to be extended. The computing tasks need more computing power, storage. The data source is not just from the branch but also from other branches. Because all the branches belong to the investment bank, the data can be shared among branches with a suitable authorization policy. Thus, it is necessary to have a data Grid in the investment bank. We can see many similar scenarios in the real world such as an auto producer with many distributed branches and so on.

To build such a data Grid, one solution is applying the way of scientific data Grid. Each branch invests to build its own computer center. Those computer centers are then connected together to form a data Grid. The users in each branch

---

* Corresponding author.

can use the Grid freely. However, such approach brings several disadvantages. It cost a great amount of money for hardware/software investment, electric power, personal and so on. The resource usage is not efficient. Usually the data mining applications are executed when all the financial data are collected. This point is usually at the end of month, end of quarter or end of year. At those time periods, all computing resources are employed and the workload is 100%. In normal time period, the workload is lower and thus many computers run wastefully or unused. There may have unfair resource usage on the Grid. Some branches may contribute little resource to the Grid but use a lot.

An alternative approach is outsourcing. This means each branch does not invest to build a computer center it self but hiring from resource providers and pay per use. In other word, the investment bank will build a data Grid over the cloud environment. This approach brings many benefits to the investment bank and its branches. Some of them are flexibility, efficiency and fair sharing.

However, up to now, there is no business model and technical solution to realize this approach. The work in this paper is the first attempt to solve this issue. In particular, the contribution of the paper includes:

- The incentive of having an economic aware data Grid. The paper will discuss the advantage of the economic aware data Grid.
- The business model. We will analyze the business relation among many business components joining the data Grid.
- The high level system architecture to realize the economic aware data Grid.

The paper is organized as follows. Section 2 describes the related works. Section 3 presents the incentives of having the economic aware data Grid and Section 4 describes the business model. The system architecture is described in section 5. Section 6 concludes the paper with a short summary.

2   Related Works

Current large-scale data grid projects include the Biomedical Informatics Research Network (BIRN) [2], the Southern California Earthquake Center (SCEC) [3], and the Real-time Observatories, Applications, and Data management Network (ROADNet) [4], all of which make use of the SDSC Storage Resource Broker as the underlying data grid technology. These applications require widely distributed access to data by many people in many places. The data grid creates virtual collaborative environments that support distributed but coordinated scientific and engineering research. The economic aspects are not considered in those projects.

The literature recorded many works related to utility computing in the Computational Grid [5,6,7,8]. Many economic principles are analyzed in the context of the utility Grid which allow resource owners, users, and other stakeholders to make value and policy statements. However all of them considered only the computation resources. The combination of many data sources to form a data Grid as described in our scenario has not been mentioned.

Related to the data Grid, in [9] a cost model for distributed and replicated data over a wide area network is presented. Cost factors for the model are the network, data server and application specific costs. Furthermore, the problem of job execution is discussed under the viewpoint of sending the job to the required data (code mobility) or sending data to a local site and executing the job locally (data mobility). However, in the model, the cost is not money but the time to perform the job execution. With this assumption, the system is pseudo economic-aware. More over, the infrastructure works with the best effort mechanism. The QoS and resource reservation are not considered. Thus, it does not suit with the business environment.

Heiser et. al. [10] proposed a commodity market of storage space within the Mungi operating system. In which, the system focuses on the extra accounting system used for backing store management. The model is designed such that all accounting can be done asynchronously to operations on storage objects, and hence without slowing down such operations. It is based on bank accounts from which rent is collected for the storage occupied by objects. Rent automatically increases as available storage runs low, forcing users to release unneeded storage. Bank accounts receive income, with a taxation system being used to prevent excessive build up of funds on underutilised accounts. However, the system considers only the storage resource and the scope of the system is just inside an organization.

Buyya in [11] discusses the possible use of economy in a scientific data Grid environment, in particular in the DataGrid project. Specifically, a token-exchange approach is proposed to regulate demand for data access from the servers of the data grid. For example, a token may correspond to 10KB of data volume. By default, a single user may only access as much data as he has tokens. This gives other users a chance to access data. The amount of data that they access for a given token needs to be based on parameters such as demand, system load, QoS requested, etc. The users can trade-off between QoS and tokens. However, the negotiation/redistribution of tokens after their expiration, their mapping to real money and the pricing policies of storage servers are not discussed in such an approach. More over, this work focuses on the resource providers level while in our work we focus on the system built up over the commercial resource providers.

3   The Incentive of Having the Economic Aware Data Grid

The economic aware data Grid brings many advantages. First of all, it is flexibility. Because of employing the cloud computing environment, each branch could have Grid resources on demand. In the critical period, they could hire more resources from the cloud. When they do not need Grid resources any more, they could return them to the cloud.

It is economic efficiency. Assume that each month, a branch needs 10.000 GB storage, 10 computers for the daily workload in 28 days, and needs 100 computers for the critical workload in 2 days. The investment of such a cluster is about 1

million Euro. Assume that the life cycle of the cluster is 5 years. Assume that the branch spends 20.000   each month including the personal cost, electricity and so on for running the cluster. Thus, the branch must pay 2,2 million Euro in 5 years. If the branch uses Cloud services, according Amazon pricing schema, the cost in 5 years is approximately 200.000.

The economic aware data Grid provides a fair resource usage. Each branch independently pays for its Grid resource usage. Thus, if a branch uses many Grid resources, it has to pay more and vice versa.

4   The Business Model Proposal

Figure 1 presents the business relationship among many components joining the system. The business model includes three main components.
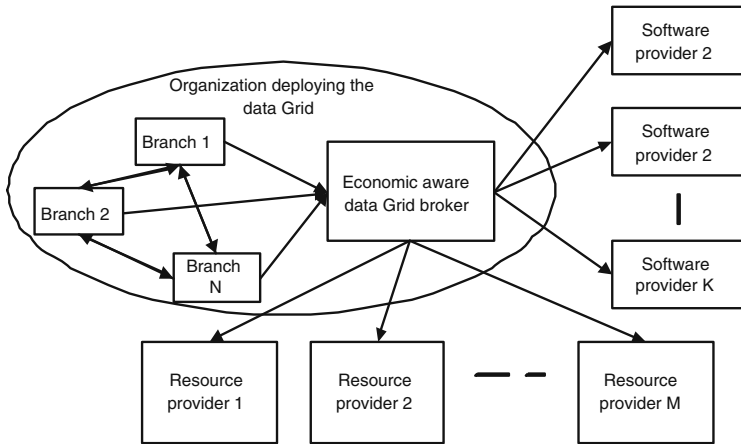

Fig. 1. The business model of the system

4.1   Resource Provider

For the utility computing business model, the resource provider provides storage capacity and computing power service. Thus, the computing power, storage, and network bandwidth are main resources to be traded. In the real environment, for example the Amazon pricing schema, the cost of network bandwidth, which is visualized through the data transfer in/out, is located inside the cost of the storage service and the computing power service. The provider should have its own accounting, charging, and billing module and metering, job deployment and so on. We assume that the price of using resource is published. To ensure the QoS, the resource provider should have advance resource reservation.

The resource provider charges the user using storage service following fees: fee of using storage, fee of having data transfer in/out, fee of having request to manipulate the data. The resource provider charges the user using computation service following fees: fee of using computing nodes, fee of having data transfer in/out.

4.2   Software Provider

Software provider is a business entity having software. The software provider provides software service. In particular, he provides software and its license so the software can work under negotiated condition. The income of the software provider is from selling software license.

4.3   Organization Deploying the Data Grid

Each branch should have its own separate data storage on the Grid because the data is very important with each branch. In particular, each branch hires basic storage service from a resource provider. The data are then shared with many authorization levels to other branches. To coordinate the working of many data sources and data exploitation activities, it is necessary to have a Grid middle ware called economic aware data Grid broker.

The economic aware data Grid broker should belong to the organization deploying the data Grid because of following reasons.

– It saves the organization many cost of using broker service.
– It is easier for organization to apply cost optimization policy when having its own control system.
– Giving data management task for a third party is not as trustable as by its self.

The economic aware data Grid broker does two main tasks. The first is the data management service. It includes data transfer service, replication service, authorization and so on. The second task is the job execution service. It receives the requirement from user, gets software, locates data, reserves computing resource, does deployment, runs the software and return the result. In fact, two tasks relate closely with each other. When the job produces output data, it must also be stored/replicated some where on the data Grid. The contract with software provider and resource provider are realized with Service Level Agreement (SLA).

The user in each branch puts/gets/finds data and runs jobs on the data Grid. Thus the branch has to pay the cost of using storage and computing power for resource provider, the cost of using software for the software providers. From the cost of using storage service, we can see that the cost includes data transfer in/out cost. Thus, if user of branch 2 does many copies from the data storage of branch 1, letting branch 1 pay for the transfer cost is unfair. Thus, it is necessary to have payment among branches to ensure that every branch is fine. They use much they have to pay much. They use less then they have to pay less.

The goal of the economic aware data Grid broker is managing the data Grid in away that minimize the cost of the organization.

5   High Level System Architecture

Figure 2 presents the high level system architecture of the economic aware data Grid. In this section, we will describe the working mechanism of the system.
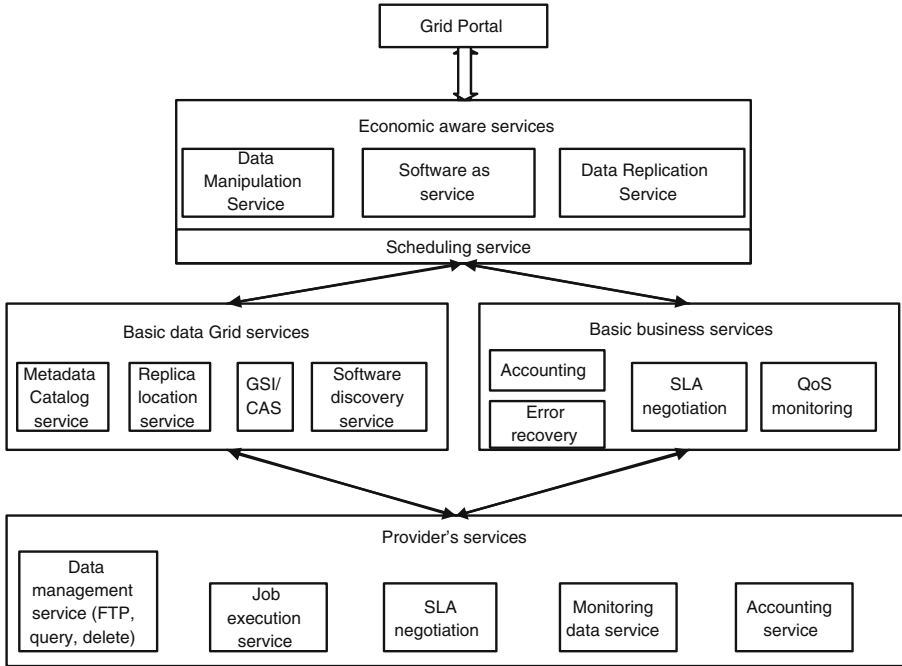
Fig. 2. High level system components

5.1   Data Manipulation

The user of a branch puts/finds/downloads/deletes a file to the Grid. As each branch has a separate storage on the Grid, the file should be put to that storage. In this case, the system will do following works:

- The Grid receives the requirement with the Grid Portal. The Grid Portal invokes the Metadata Catalog Service to find the appropriate information. If the request is put, the system returns the data storage. If the request is find/query/delete, the system returns the data location.
- The Grid Portal invokes services provided by the service provider to handle the request.
- When the request completes or fails, the Grid Portal notifies to the user. If success, the Grid Portal stores data information to Metadata Catalog Service, Replica Location service, stores the accounting information to the accounting service.

5.2   Replication

The request for replication appears in many cases: The system needs to replicate a file on Grid following a defined replication plan defined; The user demands a file

replication; A job has a request for a particular file in the Grid; The replication can also be triggered by analyzing the patterns of previous file requests. Thus, there is the migration of files toward sites that show a corresponding increased frequency of file access requests. When having the replication request, the system will do following works.

- The Data Replication service receives request, read and interpret it.
- The Data Replication service invokes scheduling service to find out a suitable replication location.
- The scheduling service discovers candidate resources, matches the user's requirements and the candidate resources in an optimal way and then returns to the Data Replication service.
- The Data Replication service reserves bandwidth with resource provider by an SLA.
- The Data Replication service invokes the FTP service of the determined resource provider for transferring data.
- The Data Replication service invokes monitoring module to monitor the QoS.
- If success, the Data Replication service stores data information to Metadata Catalog Service, Replica Location service, stores the accounting information to the accounting service.

5.3   Job Execution

When the user wants to run a job, he provides the information of the job such as software name, name of the input/output data, resource requirement, deadline and so on. The system will do following works.

- The Grid Portal invokes the SaaS service.
- The SaaS invokes the Software Discovery service to find the location of the software provider.
- The SaaS invokes the Metadata Catalog Service, Replica Location service to find the location of data file.
- The SaaS invokes Scheduling service to find the suitable resource provider.
- The SaaS signs SLA of hiring software, computing resource, bandwidth with software provider, resource provider.
- The SaaS download the software and data to the execution site and execute the job.
- During the execution, monitoring module is invokes to observe the QoS.
- During the execution, if having some error, SaaS will revoke the error recovery.
- When the execution finishes, the SaaS move the output data to the defined placed and update Metadata Catalog Service, Replica Location service.
- The SaaS store accounting information to the account service.

It is noted that unlike the general SaaS, the number of software used in the system is not so big. Thus we need only a simple Software discovery service.

6   Conclusion

This paper has presented the incentive, the business model and the required technical infrastructure of the economic aware data Grid. The economic data aware Grid has many benefits comparing to the traditional form of the data Grid such as flexibility, efficiency and fair sharing. The business relation among business entities is clearly defined to gain the fairness. The high level architecture and is described to have an overview of the working mechanism of the system. This is the first step in realizing the whole system.

References

1. Groep, D., Templon, J., Loomis, C.: Crunching real data on the Grid: practice and experience with the European DataGrid. Concurrency Computat.: Pract. Exp. 18, 925–940 (2006)
2. Peltier, S.T., Ellisman, M.H.: The Biomedical Informatics Research Network, The Grid, Blueprint for a New Computing Infrastructure, 2nd edn. Elsevier, Amsterdam (2003)
3. Southern California Earthquake Data Center team: Seismogram Transfer Program (STP) Reference Manual,
   `http://www.data.scec.org/STP/STP_Manual_v1.01.pdf`
4. Fricke, T., Ludaescher, B., Altintas, I., Lindquist, K., Hansen, T., Rajasekar, A., Vernon, F., Orcutt, J.: Integration of Kepler with ROADNet: Visual Dataflow Design with Real-time Geophysical Data. Poster in American Geophysical Union Fall Meeting (2004)
5. Treaster, M., Kiyanclar, N., Koenig, G.A., Yurcik, W.: A Distributed Economics-based Infrastructure for Utility Computing, ArXiv Computer Science e-prints (2004)
6. Brunelle, J., Hurst, P., Huth, J., Kang, L., Ng, C., Parkes, D.C., Seltzer, M., Shank, J., Youssef, S.: EGG: an extensible and economics-inspired open grid computing platform. In: Proceeding of the 3rd International Workshop on Grid Economics and Business Models, GECON 2006, pp. 140–150 (2006)
7. Wolski, R., Brevik, J., Plank, J.S., Bryan, T.: Grid Resource Allocation and Control Using Computational Economies. In: Berman (ed.) Grid Computing: Making The Global Infrastructure a Reality. John Wiley & Sons, Chichester (2003)
8. Buyya, R.: Economic-based Distributed Resource Management and Scheduling for Grid Computing, PhD thesis at School of Computer Science and Software Engineering Monash University, Melbourne, Australia (2002)
9. Stockinger, H., Stockinger, K., Schikuta, E., Willers, I.: Towards a Cost Model for Distributed and Replicated Data Stores. In: Proceedings of the 9th Euromicro Workshop on Parallel and Distributed Processing (PDP 2001), pp. 461–467 (2001)
10. Heiser, G., Elphinstone, K., Vochteloo, J., Russell, S., Liedtke, J.: The Mungi single-address-space operating system. Software–Practice & Experience 29(9), 901–928 (1998)
11. Buyya, R., Abramson, D., Giddy, J., Stockinger, H.: Economic models for resource management and scheduling in Grid computing. Concurrency Computat.: Pract. Exper. 14, 1507–1542 (2002)